

# DNA Damage in Plant Herbarium Tissue

Martijn Staats<sup>1</sup>, Argelia Cuenca<sup>2</sup>, James E. Richardson<sup>3,4</sup>, Ria Vrieling-van Ginkel<sup>1</sup>, Gitte Petersen<sup>2</sup>, Ole Seberg<sup>2</sup>, Freek T. Bakker<sup>1\*</sup>

**1** Biosystematics Group, Wageningen University, Wageningen, The Netherlands, **2** Laboratory of Molecular Systematics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, **3** Tropical Diversity Section, Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom, **4** Laboratorio de Botánica y Sistemática, Universidad de Los Andes, Bogotá, Colombia

## Abstract

Dried plant herbarium specimens are potentially a valuable source of DNA. Efforts to obtain genetic information from this source are often hindered by an inability to obtain amplifiable DNA as herbarium DNA is typically highly degraded. DNA post-mortem damage may not only reduce the number of amplifiable template molecules, but may also lead to the generation of erroneous sequence information. A qualitative and quantitative assessment of DNA post-mortem damage is essential to determine the accuracy of molecular data from herbarium specimens. In this study we present an assessment of DNA damage as miscoding lesions in herbarium specimens using 454-sequencing of amplicons derived from plastid, mitochondrial, and nuclear DNA. In addition, we assess DNA degradation as a result of strand breaks and other types of polymerase non-bypassable damage by quantitative real-time PCR. Comparing four pairs of fresh and herbarium specimens of the same individuals we quantitatively assess post-mortem DNA damage, directly after specimen preparation, as well as after long-term herbarium storage. After specimen preparation we estimate the proportion of gene copy numbers of plastid, mitochondrial, and nuclear DNA to be 2.4–3.8% of fresh control DNA and 1.0–1.3% after long-term herbarium storage, indicating that nearly all DNA damage occurs on specimen preparation. In addition, there is no evidence of preferential degradation of organelle versus nuclear genomes. Increased levels of C→T/G→A transitions were observed in old herbarium plastid DNA, representing 21.8% of observed miscoding lesions. We interpret this type of post-mortem DNA damage-derived modification to have arisen from the hydrolytic deamination of cytosine during long-term herbarium storage. Our results suggest that reliable sequence data can be obtained from herbarium specimens.

**Citation:** Staats M, Cuenca A, Richardson JE, Vrieling-van Ginkel R, Petersen G, et al. (2011) DNA Damage in Plant Herbarium Tissue. PLoS ONE 6(12): e28448. doi:10.1371/journal.pone.0028448

**Editor:** Carles Lalueza-Fox, Institut de Biologia Evolutiva - Universitat Pompeu Fabra, Spain

**Received:** August 25, 2011; **Accepted:** November 8, 2011; **Published:** December 5, 2011

**Copyright:** © 2011 Staats et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by SYNTHESIS2 made available by the European Community Research Infrastructure under FP7 [‘Synthesis of Systematic Resources’, 226506-CP-CSA-Infra]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: freek.bakker@wur.nl

## Introduction

The world’s approximately 3400 herbaria (<http://sciweb.nybg.org/science2/IndexHerbariorum.asp>) contain an immense number of plant specimens covering virtually all known species, making herbaria not only invaluable assets for understanding plant biodiversity [1,2], but also largely underutilised genomic treasure troves. The expansion of next-generation sequencing (NGS) capabilities will potentially open up possibilities for cost-effective sequencing of genomes from type specimens and rare or extinct species stored in herbaria [3]. Although, DNA extraction results in irreparable damage to specimens, which conflicts with their historic and scientific importance, typically only a few milligrams of herbarium material need to be sampled. Nevertheless for small herbarium specimens (e.g. some Brassicaceae) or type specimens this can be too much, as the whole specimen basically has to be sacrificed. Therefore, considerable effort has been spent on optimizing DNA extraction protocols [4–6]. Furthermore, herbarium DNA is typically highly degraded into low molecular weight fragments [7–9]. Up until twenty years ago, herbarium specimen preparation techniques were not aimed at preserving DNA. Thus commonly used collection methods involved chemical treatments of specimens with formalin or ethanol, both of which severely affect DNA preservation in plants [7,10,11].

The occurrence of apuric sites, deaminated cytosine residues, and oxidized guanine residues are the main types of damage known from studies *in vivo* and on ancient DNA [12,13]. In living cells, such sites can have lethal consequences and are efficiently repaired [14]. Herbarium specimen preparation, however, induces high levels of metabolic and cellular stress responses and ultimately cell death resulting in irreparably damaged DNA [15]. The post-mortem DNA damage inflicted during specimen preparation may be higher in organelles, as they are the major source of reactive oxygen species (ROS) known to inflict oxidative nucleotide damage [16,17]. Once preserved, specimens in all major herbaria are normally (but not continuously) protected from the damaging effects of ultraviolet light and stored at moderate temperatures and at relatively low humidity, and often subjected to a two-yearly –20°C freezing cycle.

Damaged nucleotides in herbarium DNA may result in damage-specific nucleotide mis-incorporations (miscoding lesions) by DNA polymerase enzymes during amplification [18,19]. In contrast to such polymerase-by-passable damage, strand-breaks and other DNA modifications block polymerases and thus prevent amplification. Qualitative and quantitative assessment of DNA post-mortem damage is therefore essential to determine the accuracy of DNA sequence data from herbarium specimens. The first aim of this study was to assess DNA damage as a result of

polymerase non-bypassable damage using quantitative real-time PCR for multiple plastid, mitochondrial, and nuclear DNA regions. Secondly, levels of miscoding lesions in herbarium DNA were assessed using 454-sequencing of amplicons derived from each of the three genomic compartments. Using fresh and herbarium specimens of up to 114 years old, taken from the same individuals, allows a quantitative assessment of post-mortem DNA damage. Post-mortem damage was assessed, i) directly after herbarium specimen preparation by comparing results from fresh tissue and young herbarium specimens, and ii) after long-term herbarium storage by comparing results from young and old (>65 yrs.) herbarium specimens. Through statistical comparison, we investigated whether polymerase misincorporation errors alone explain the levels of miscoding lesions observed or whether they represent true damage-derived lesions in herbarium DNA. Finally, levels of DNA damage in the plastid, mitochondrial, and nuclear genomic compartments were compared in order to test for preferential DNA degradation among them.

## Results

### Degree of DNA fragmentation caused by herbarium preservation

The oldest herbarium material used was a 114 year old specimen of *Liriodendron tulipifera*. The other specimens included were *Ginkgo biloba* (107 years old) and *Laburnum anagyroides* (65 years old). The DNA extracted from these specimens was typically highly degraded and DNA fragment sizes were mostly below 1 kb (Figure S1). DNA extracts from young herbarium specimens (i.e., dated 8 July 2010) dried for 18 hours at 60°C did not contain high molecular weight DNA, and the average fragment size was below 10 kb. DNA extracts prior to filter cleanup were brownish, whereas cleanup yielded clear extracts with  $A_{260}/A_{280}$  ratios between 1.7 and 2.0.

DNA yields from old and young herbarium specimens were not significantly different ( $P=1.000$ ; Table S6) and ranged between 33.91 and 103.40 ng DNA/mg tissue (Table S1). DNA yields from fresh tissue were on average 4.4–5.4 times higher, which may reflect differences in fresh versus herbarium DNA extraction efficiencies as low molecular weight DNA (DNA molecules <100 bp and nucleotides) is known to be less-efficiently recovered.

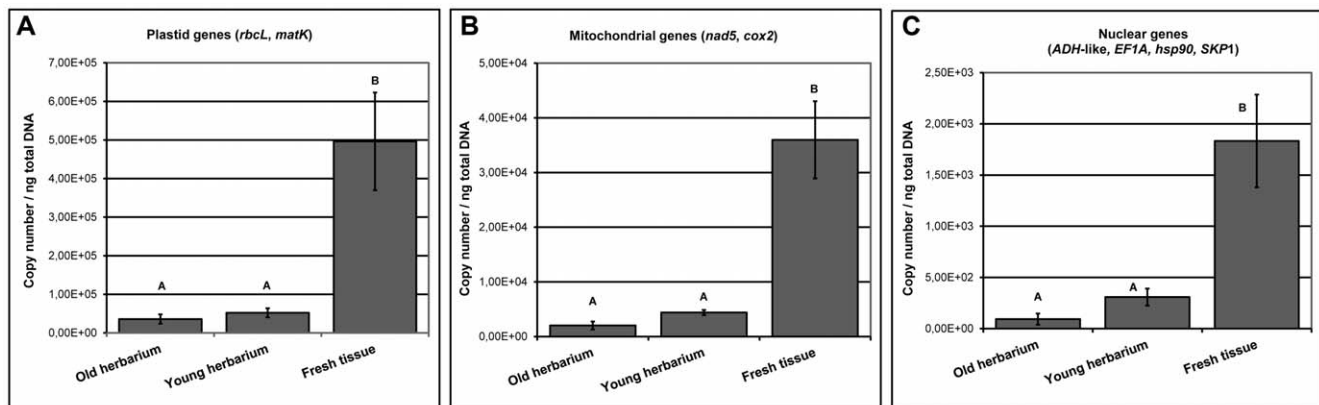
Any potential damage to herbarium DNA is likely to affect amplification efficiencies and calculated threshold cycle (Ct)-values. Depending on the kind and extent of damage this may lead

to erroneous detection of gene copy numbers in herbarium DNA due to PCR jumping artefacts [28]. In this study, however, this effect was likely to be limited, because the qPCR amplicon sizes (80–140 bp) were always shorter than maximum fragments sizes ( $\leq 1$  kb) in herbarium extracts. Furthermore, no amplification inhibition was observed in tests with serially diluted herbarium DNA samples (results not shown). Amplicon copy numbers for all plastid, mitochondrial, and nuclear gene regions in fresh and herbarium tissues are presented in Table S5. Mean gene copy numbers were measured by calculating the mean of plastid, mitochondrial or nuclear amplicon numbers across genes per genomic compartment (Table S6).

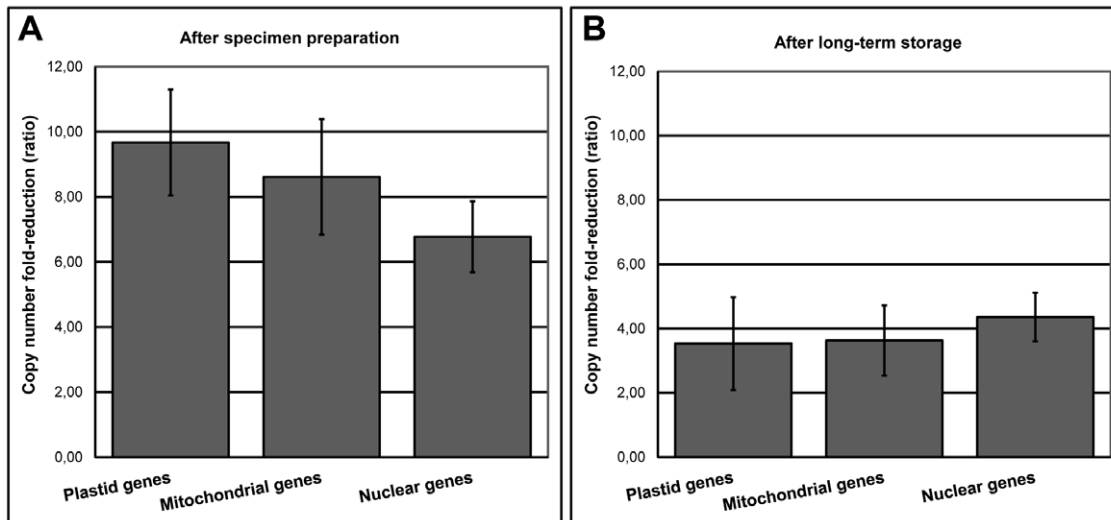
Fresh leaf tissue yielded significantly higher plastid, mitochondrial and nuclear gene copy numbers than young and old herbarium specimens (Figure 1A–C). As measured by the fresh tissue/young herbarium gene-copy ratio, plastid copy numbers after specimen preparation were on average 9.7-fold reduced (Figure 2A), which equates to a mean reduction of 444196 plastid DNA copies (per ng of total DNA) after specimen preparation (Table S6). Compared to young herbarium specimens, plastid copy numbers were on average 3.5-fold reduced following long-term storage (young/old herbarium ratio; Figure 2B), which equates to a mean reduction of 16322 plastid DNA copies after long-term storage. Likewise, mitochondrial DNA was reduced with 31589 copies after specimen preparation and 2373 copies after long-term storage, representing an mtDNA copy number reduction of 8.6 and 3.63-fold, respectively (Figure 2).

Due to the inability to produce optimal qPCR standard curves for some nuclear gene amplicons (results not shown), different sets of nuclear genes were assayed for the four species. Since the assayed loci were all predicted to be low-copy nuclear genes based on PLAZA 2.0 BLAST searches (Table S2), we believe it justifiable to compare copy numbers for different nuclear loci. Nuclear DNA was reduced with 1524 copies after specimen preparation and 215 copies after long-term storage, representing a nDNA copy number reduction of 6.8 and 4.3-fold, respectively (Figure 2).

Our results indicate that there is no difference in the degree of DNA fragmentation between plastid, mitochondrial, and nuclear genomes in herbarium specimens, as the reduction in nDNA copy numbers is not statistically different from those in plastid or mitochondrial DNA (Figure 2). Organelle DNA therefore does not appear to be preferentially degraded in herbarium tissue both directly following drying, or after long-term storage. Copy number reduction directly after drying was more pronounced than that



**Figure 1. Copy numbers for herbarium specimens and fresh tissues for: (A) plastid genes (B) mitochondrial genes (C) nuclear genes.** Values statistically different at 5% significance level in post-hoc tests are indicated by different letters (A or B). doi:10.1371/journal.pone.0028448.g001



**Figure 2. Copy number fold-reduction of plastid, mitochondrial and nuclear genes in herbarium specimens (A) after specimen preparation (fresh tissue/young herbarium ratio), and (B) after long-term herbarium storage (young/old herbarium ratio).** Values were not statistically different at 5% significance level. doi:10.1371/journal.pone.0028448.g002

following long-term storage. We estimate that, depending on the genomic compartment, 10.5–16.8% of the total DNA extracted from young herbarium specimens were amplifiable using *Taq* polymerase (Table 1). After correction for losses in DNA yield due to herbarium specimen preparation, however, we estimate the proportion of amplifiable copy numbers for plastid, mitochondrial, and nuclear DNA to be 2.4–3.8% of that of fresh controls and 1.0–1.3% after long-term storage.

#### Elevated levels of C→T/G→A transitions in old herbarium plastid DNA

The average depth of coverage per amplicon was 4640×, and therefore the level of detection was sufficient to detect 1–5% variation of single base changes (Roche application note No. 5). The overall nucleotide misincorporation error rate recorded in fresh leaf tissue was  $\sim 1.4 \times 10^{-3}$  per nucleotide (44127 total substitutions/31933585 total nucleotides in fresh tissue; Table 2). This background substitution level in DNA extracted from fresh leaf tissue was assumed not only due to nucleotide misincorporations that arise from DNA polymerase errors, but also because of 454-sequencing errors and potential damage that may have arisen during DNA extraction.

Chi-square ( $\chi^2$ ) tests of independence were used to compare the observed distributions of substitutions in herbarium DNA to the

expected distributions in fresh leaves. Because the amplicons were generated by PCR, the actual strand of origin of potential miscoding lesions cannot be identified. Therefore, the data were summarized into six complementary pairs of nucleotide substitution.  $\chi^2$  testing of the plastid data provided strong support that the quantitative distributions of substitutions summarized over the six substitution types in young herbarium specimens ( $\chi^2 = 209.37$ ;  $df = 5$ ;  $P < 0.001$ ) and old ones ( $\chi^2 = 4539.16$ ;  $df = 5$ ;  $P < 0.001$ ) were significantly different from those in fresh tissue (Table 2). Similarly high chi-square values were observed for mitochondrial and nuclear data (Table 2).

Analyses of variance (ANOVA) were performed for each of the six substitution types separately (Table S7). This enabled the identification of specific substitution types in herbarium specimens that were over-represented and thus may be attributed to DNA post-mortem damage. C→T/G→A transitions in plastid DNA from old herbarium specimens occur significantly more frequently than in plastid DNA from fresh and young herbarium ( $F = 37.42$ ;  $P < 0.001$ ; Table S7). No increased levels of C→T/G→A transitions were detected in herbarium mitochondrial and nuclear DNA. Plastid C→T/G→A transitions constitute 21.4% (4812 out of 22527 substitutions) of the observed miscoding lesions in old herbarium plastid DNA (Table 2). Although, the overall percentage of total observed nucleotide sites at which the

**Table 1. Mean DNA yield, mean gene copy numbers, and DNA yield loss-corrected copy numbers for plastid, mitochondrial and nuclear DNA relative to fresh control.**

Sample type	% mean DNA yield <sup>a</sup>	% mean gene copy numbers <sup>b</sup>		
		Plastid	Mitochondrial	Nuclear
Fresh tissue	100	100	100	100
Young herbarium	22.6	10.5 (2.4)	12.2 (2.8)	16.8 (3.8)
Old herbarium	18.7	7.2 (1.3)	5.6 (1.0)	5.1 (1.0)

<sup>a</sup>Percentages were calculated relative to fresh controls, which were set at 100%. Mean DNA yields and mean gene copy numbers were taken from Table S6.

<sup>b</sup>Percentages of mean gene copy numbers corrected for DNA yield losses relative to fresh control (in brackets) are calculated as: (% mean DNA yield) × (% mean gene copy numbers).

doi:10.1371/journal.pone.0028448.t001

**Table 2.** Chi-square ( $\chi^2$ ) tests of independence (at  $p \geq 0.05$  level) on number of nucleotide substitutions observed in fresh and herbarium DNA.

DNA type	Substitution type	Fresh tissue			Young herbarium			Old herbarium			
		Observed (%) <sup>b</sup>	Observed (%)	Observed, corrected <sup>c</sup>	Observed (%)	Observed (%)	Observed, corrected <sup>c</sup>	Observed (%)	Observed (%)	Observed, corrected <sup>c</sup>	
<b>Plastid</b>	<b>Total nucleotides<sup>a</sup></b>	10746437	14193327	10746437	14193327	10746437	14101498	10746437	14101498	10746437	
	<b>Total substitutions</b>	11298	16047	12149.94	12149.94	12149.94	22527	17167.32	22527	17167.32	
	(A→G/T→C)	8504 (0.0791)	11634 (0.0817)	8808.65	8808.65	10.91	15040 (0.1066)	11461.65	15040 (0.1066)	11461.65	
	(A→C/T→G)	264 (0.0024)	427 (0.0030)	323.30	323.30	13.32	619 (0.0044)	471.72	619 (0.0044)	471.72	
	(A→T/T→A)	521 (0.0048)	585 (0.0041)	442.93	442.93	11.69	937 (0.0066)	714.07	937 (0.0066)	714.07	
	(C→A/G→T)	272 (0.0025)	386 (0.0027)	292.26	292.26	1.51	1021 (0.0072)	778.08	1021 (0.0072)	778.08	
	(C→G/G→C)	49 (0.0046)	79 (0.0006)	59.82	59.82	2.39	98 (0.0007)	74.68	98 (0.0007)	74.68	
	(C→T/G→A)	1688 (0.0157)	2936 (0.0207)	2222.98	2222.98	169.55	4812 (0.0341)	3667.12	4812 (0.0341)	3667.12	
					209.37 (<0.001)						4539.16 (<0.001)
	<b>Mitochondrial</b>	<b>Total nucleotides</b>	8566629	9769969	8566629	9769969	8566629	8738228	8566629	8738228	8566629
	<b>Total substitutions</b>	14050	16174	14181.89	14181.89	16937.72	17277	16937.72	17277	16937.72	
	(A→G/T→C)	7580 (0.0885)	8828 (0.090)	7740.68	7740.68	3.41	8113 (0.0928)	7953.68	8113 (0.0928)	7953.68	
	(A→C/T→G)	380 (0.0044)	408 (0.0042)	357.75	357.75	1.30	525 (0.0060)	514.69	525 (0.0060)	514.69	
	(A→T/T→A)	317 (0.0037)	403 (0.0041)	353.36	353.36	4.17	328 (0.0038)	321.56	328 (0.0038)	321.56	
	(C→A/G→T)	569 (0.0066)	634 (0.0065)	555.91	555.91	0.30	834 (0.0095)	817.62	834 (0.0095)	817.62	
	(C→G/G→C)	198 (0.0023)	247 (0.0025)	216.58	216.58	1.74	281 (0.0032)	275.48	281 (0.0032)	275.48	
	(C→T/G→A)	5006 (0.0584)	5654 (0.0579)	4957.61	4957.61	0.47	7196 (0.0823)	7054.69	7196 (0.0823)	7054.69	
				11.39 (0.044)						1043.60 (<0.001)	
<b>Nuclear</b>	<b>Total nucleotides</b>	12620519	13953891	12620519	13953891	12620519	8608206	12620519	8608206	12620519	
	<b>Total substitutions</b>	18779	20334	18390.97	18390.97	31182.55	21269	31182.55	21269	31182.55	
	(A→G/T→C)	13084 (0.1036)	13923 (0.0997)	12592.58	12592.58	18.46	11726 (0.1362)	17191.53	11726 (0.1362)	17191.53	
	(A→C/T→G)	547 (0.0043)	595 (0.0043)	538.14	538.14	0.14	826 (0.0096)	1211.00	826 (0.0096)	1211.00	
	(A→T/T→A)	682 (0.0054)	702 (0.0050)	634.92	634.92	3.25	1380 (0.0160)	2023.22	1380 (0.0160)	2023.22	
	(C→A/G→T)	473 (0.0037)	498 (0.0036)	450.41	450.41	1.08	681 (0.0079)	998.42	681 (0.0079)	998.42	
	(C→G/G→C)	243 (0.0019)	348 (0.0025)	314.75	314.75	21.18	417 (0.0048)	611.37	417 (0.0048)	611.37	
	(C→T/G→A)	3750 (0.0297)	4268 (0.0306)	3860.17	3860.17	3.24	6239 (0.0725)	9147.02	6239 (0.0725)	9147.02	
					47.35 (<0.001)						31642.64 (<0.001)

<sup>a</sup>Total numbers of adenine, thymine, cytosine and guanine in the dataset.<sup>b</sup>Numbers of observed substitutions and their percentages of total nucleotides observed (in brackets).<sup>c</sup>The numbers of observed substitutions were corrected for differences in total sequence data (nucleotides) between fresh and herbarium data.Separate analyses were performed for plastid, mitochondrial, and nuclear DNA. Observed numbers in the fresh tissue data were used as expected numbers in  $\chi^2$ -tests.

doi:10.1371/journal.pone.0028448.t002

C→T/G→A transitions occurred is low (0.0341%), the C→T/G→A rate was approximately twice as high as in fresh tissue (0.0157%; Table 2). Based on our results therefore, the estimated C→T/G→A rate in herbarium plastid DNA during storage can be calculated to be  $1.53 \times 10^{-6}$  per nucleotide per year (Table 3). Furthermore, A→C/T→G transversions in nuclear DNA occurred significantly more frequently in old herbarium than in fresh material and young herbarium DNA ( $F=10.74$ ;  $P=0.005$ ; Table S7). However, they constitute only 3.9% (826 out of 21269 substitutions) of the total miscoding lesions, and appear to play little or no role in damage-derived miscoding lesions in herbarium DNA (Table 2).

## Discussion

Herbaria are major depositories for conserved plant material, and their combined collections provide an unparalleled record of the world's flora [1,29]. It is not surprising, therefore, that herbarium specimens are a common source of DNA for studies on plant evolution. Despite this, little research has focused on post-mortem DNA damage in herbarium material and its influence on data quality. As far as we know, this is the first study to quantify the prevalent damage types in herbarium DNA following an experimental approach.

### Degree of DNA fragmentation caused by herbarium DNA preservation

Our results show that the most obvious form of post-mortem damage in herbarium DNA is double-stranded breaks. It is likely that the high temperatures (60–70°C) at which herbarium specimens are typically dried today causes cells to rupture quickly, concomitantly releasing nucleases, ROS, and other cellular enzymes. Such physiological conditions resemble necrosis, and this overwhelming cellular stress typically causes DNA to be degraded randomly into smaller fragments, appearing as a smear on agarose gels [30,31]. Our experience is that DNA extracts from herbarium tissues rarely show oligo-nucleosomal DNA fragmentation [30], which is an indicator for programmed cell death (PCD). PCD is a highly coordinated death process, which is known to activate extra nucleases and ROS that damage DNA [30,32]. Signs of DNA 'laddering' may indicate that the plant material was preserved too slowly and that the plant had experienced abiotic stresses for prolonged periods (hours, days) during preparation. Rapid desiccation of plant material has been shown to be the best way to preserve plant DNA [10], as it limits the PCD damage process.

Given that mtDNA and plastid DNA are close to major sites of ROS production and have no histones and no chromatin structure, one might expect higher levels of damage in organelle DNA during herbarium specimen preparation. However, no

preferential degradation of mtDNA or plastid DNA was observed, and our results, therefore, indicate that DNA damage levels in each of the three genomic compartments are equal. The high temperatures at which herbarium specimens are typically dried most likely directly affects the degree to which DNA is preserved. Heating is known to greatly accelerate hydrolytic depurination, which in combination with the spontaneous cleavage of the phosphodiester backbone by  $\beta$ -elimination will result in strand breaks [33,34].

Our results indicate that up to 89.5% of DNA of young herbarium specimens may not be accessible to *Taq* polymerase (100% minus 10.5% mean plastid copy numbers; Table 1). The poor amplification success was not due to the presence of PCR inhibitors, as herbarium extracts were not found to delay amplification of an exogenous DNA control. Another possible explanation is that damage in herbarium DNA may be locus- or region-specific, and there is possibly large copy number heterogeneity in herbarium DNA samples. Gene copy numbers within each genomic compartment, however, were found to be of the same order of magnitude, rejecting a region-specific damage hypothesis. Admittedly, the representation of each genomic compartment by only two or three loci is a coarse representation of (the complexity of) organellar and nuclear genomes. A more representative sampling will be needed in order to exclude potential hotspots of degradation in herbarium DNA. A further consideration is that herbarium DNA was severely modified, not only by double stranded breaks, but also by inter-strand cross-links, abasic sites, or other structural modifications. These types of polymerase non-bypassable damage will also prevent DNA molecules from being amplified and sequenced. As an example, some studies on ancient DNA have concluded that DNA molecules were highly modified by blocking lesions [35,36]. It will be of great value to assess the levels of abasic sites and other types of polymerase non-bypassable damage in herbarium DNA, as this will facilitate future DNA repair and rescuing strategies. Such strategies can help to improve the retrieval of accurate sequence information from herbarium specimens. For example, by pre-treating the herbarium DNA with enzymes involved in base excision repair, specific types of blocking lesions could be removed [37]. Furthermore, the use of engineered polymerases capable of extending beyond blocking lesions may allow for the amplification of highly damaged herbarium DNA molecules [38].

### Degree of DNA sequence modification in herbarium DNA

Increased levels of C→T/G→A transitions were observed in plastid DNA of old herbarium specimens as compared to young ones and fresh tissues. Therefore, polymerase and sequencing errors alone cannot account for all of the observed nucleotide substitutions (as we would expect these to occur without bias), which may be better explained by added misincorporations

**Table 3.** Estimated (C→T/G→A) rate in herbarium plastid DNA during herbarium storage.

Species	Number of substitutions per 10 <sup>6</sup> nucleotides			(C→T/G→A) per nucleotide per year ± SD
	Young herbarium	Old herbarium	Old herbarium age (yrs.)	
<i>Laburnum</i>	197.55	330.02	65	$2.04 \times 10^{-6}$
<i>Ginkgo</i>	229.09	384.02	107	$1.45 \times 10^{-6}$
<i>Liriodendron</i>	186.14	313.52	114	$1.12 \times 10^{-6}$
Average				$1.53 \times 10^{-6} \pm 4.66 \times 10^{-7}$

doi:10.1371/journal.pone.0028448.t003

resulting from miscoding lesions present in old herbarium DNA molecules. The observed levels were, however, low (21.4% of miscoding lesions), and we found the C→T/G→A rate to be  $1.53 \times 10^{-6}$  per nucleotide per herbarium storage year. C→T/G→A (type 2) transitions are the most common form of miscoding lesions in ancient DNA [39,40] and are ascribed to hydrolytic deamination of cytosine to uracil [18,41]. Our results therefore indicate that even though herbarium specimens are protected from the greatest threats to their long-term maintenance, herbarium DNA remains susceptible to hydrolytic activity. It seems probable that moisture in the air may allow for some rehydration of DNA that will lead to hydrolytic DNA damage.

We have no comprehensive explanation for the fact that the observed increased levels of C→T/G→A transitions seem to be exclusive for herbarium plastid DNA. A possible explanation is that DNA damage is probably more easily detected in plastid genes, simply because of the far higher template copy numbers compared with mitochondrial and (low-copy) nuclear genes. To what extent GC-content could play a role is not clear: %GC is usually low in angiosperm plastid genomes [42], which would seem to contradict increased type 2 transition rate hypothesis.

Further evidence for DNA damage following long-term herbarium storage is obtained from the apparent higher levels of DNA fragmentation and associated gene copy number decrease in old (65–110 yrs.) herbarium versus young herbarium specimens (Table 1; Figure S1). It should be noted, however, that the levels of DNA damage in young and old herbarium specimens may not be directly compared. First, there is no record of the exact method that had been used to prepare the old herbarium specimens. For instance, the precise temperature at which they were dried is unknown. Moreover, the temperature of commonly used gas-ovens in the early 1900 s at the Leiden herbarium would have been difficult to control. Second, we did not account for potential seasonal and year-to-year fluctuations in DNA copy numbers in old herbarium specimens (although we collected our living specimens roughly at the same original date). Plastid DNA copy numbers are known to decline during the growing season, which is typical for maturing plant leaves [43]. Therefore, a precise quantification of DNA damage after long-term herbarium storage cannot be given. However, the present results clearly indicate that nearly all DNA damage occurred during specimen desiccation, and we estimate that only a small proportion of damage can be attributed to long-term storage (Table 1; Figure 1).

Our findings may not be directly applicable to herbarium material in general, as a large variety of desiccation methods have been used to preserve specimens. Only since the 1990 s was the routine established in herbaria to collect separate DNA samples in silica gel [10], along with the herbarium specimens. Common practice for field preparation, especially in the tropics, would have been the temporary fixation with methylated spirits (the ‘Schweinfurt method’) or 30% formaldehyde to prevent specimens from moulding. Unfortunately, use of these chemicals is known to have destructive effects on DNA [8]. In addition, there are numerous other techniques for field drying, including the use of kerosene stoves, 100-watt light-bulbs, and air-drying on a moving vehicle. Assessing implications of each method on DNA quality will not be straightforward, as the information on the exact drying method is typically not recorded. Moreover, factors that may influence the degree of DNA preservation are likely to be species-specific (e.g., contents of secondary metabolites, etc.) and relate to the physiological state of the plant when collected, and they are therefore difficult to predict [15].

## Implications

Our study confirms that herbaria are incredibly rich sources for reliable DNA sequence data. Various next-generation sequencing approaches can now be applied to severely fragmented DNA, and although more difficult it is still possible to generate DNA sequence data from them [44]. Nevertheless, our observation that polymerase-accessible DNA may be reduced by up to ~90% could cause challenges in sample preparation using current NGS approaches, especially when limited amounts of herbarium material are available. Probably a more serious issue would be the degree of sequence modification. Our results indicate that C→T/G→A substitutions may theoretically cause an incorrect DNA sequence to be produced; however, we predict the sequence error rate to be negligible (~0.03%) even if PCR products (i.e., the 750-bp *rbcL* barcode region sequence) had been sequenced from a single clone. However, the importance of these problems depends upon the type of investigation, e.g. random errors are like random noise unlikely to produce a phylogenetic signal. Indeed, so far there is no indication that, within the context of comparative studies including both fresh and old herbarium accessions, the latter share additional type 2 transitions.

The adoption of silica-gel drying of specimens generally yields higher quality DNA than most herbarium specimen preparation methods [10], and we recommend that all future collections stored for subsequent genetic analysis should continue to use this approach. However, our results confirm that herbarium DNA is a readily available resource that will be invaluable for future phylogenetic and genomic studies not least in view of the current biodiversity crisis. Other methods, e.g., proteomics or RNA studies, may require alternative conservation methods.

In a further study we will focus on elucidating the role of blocking lesions that may prevent sequencing of herbarium DNA. A study on this matter, using massive parallel sequencing of herbarium DNA, will be published elsewhere.

## Materials and Methods

### Specimen sampling and herbarium specimen preparation

Plant material (leaves) of living trees were sampled from the Leiden Botanical Garden and associated herbarium specimens (65–114 years old) of the same individuals were obtained from the collections of the National Herbarium of the Netherlands at Leiden (Table S1). All necessary permissions for the described plant and specimen sampling were obtained from the respective curators, i.e. Dr. Paul J.A. Keßler (Hortus Leiden) and Dr. Jan de Koning (Herbarium Leiden). By selecting fresh and herbarium material from the same individual we avoided possible intraspecific genetic variation (e.g., genotypic variation and variations in the amounts of cpDNA) although we acknowledge that some level of somatic variation may exist. *Ginkgo biloba* L. (Ginkgoaceae), *Laburnum anagyroides* Medik. (Fabaceae), *Liriodendron tulipifera* L. (Magnoliaceae), and *Lonicera maackii* (Rupr.) Herder (Caprifoliaceae) were sampled. Fresh leaves were sampled on 8<sup>th</sup> July 2010 and stored at  $-80^{\circ}\text{C}$  until DNA extraction. Herbarium specimens of fresh leaves from the same four species (i.e. dating 8 July 2010) were prepared by aluminium corrugate-drying [9] and are referred to hereafter as ‘young herbarium specimens’. In order to simulate common practice in herbaria, and hence to make our results representative for typical herbarium specimens, stacked samples were pressed and dried overnight in an electric oven (Binder, type IP20) at  $60^{\circ}\text{C}$  for 18 hours. As far as we could reconstruct, our old herbarium specimens had been preserved

using a similar corrugate-drying method and at similar temperature.

### DNA extraction and purification

Total genomic DNA was extracted using a modified cetyltrimethyl-ammonium-bromide (CTAB) method [20], probably the most commonly used plant DNA extraction method. Briefly, we used the following modifications: 50 mg of leaf material (with large veins removed) was weighed and placed into a 2 ml reaction tube containing five glass beads ( $\text{\O}$  3 mm). The reaction tube was submerged in liquid nitrogen, and the sample was then homogenized using beads (Retch, type MM2) for 30 seconds at 80 rpm. Liquid nitrogen freezing and homogenization were repeated until the leaf material had turned into a fine powder. One ml of CTAB buffer (2% CTAB, 2% PVP-40, 100 mM Tris-HCl, 1.4 M NaCl, 20 mM EDTA) and 12  $\mu\text{l}$   $\beta$ -mercaptoethanol were added with subsequent incubation for 60 min at 55°C. One ml of 24:1 chloroform:isoamylalcohol was then added, vortexed and centrifuged at 14,000 rpm for 4 min. The supernatant was removed, and the chloroform extraction was repeated. DNA was precipitated using 70% isopropanol at  $-20^{\circ}\text{C}$  for 2 weeks, after which the DNA was pelleted at 14,000 rpm for 5 min. DNA was then re-suspended in TE buffer and treated with RNase (Qiagen). DNA purification was performed using the Wizard DNA clean-up system (Promega Corp.) in combination with a vacuum manifold (Promega Corp.). The DNA was dissolved in 75  $\mu\text{l}$  of pre-heated elution buffer (Qiagen). DNA extractions were visualized on 1% agarose gels (Figure S1), and the quantity was measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific). DNA extractions were performed in duplicate or triplicate, and DNA yield per dry weight tissue was determined.

### Amplification and sequencing of plastid, mitochondrial and nuclear target genes

Multiple target regions were selected for primer design, of which Expressed Sequence Tag (EST) data and/or other (partial) sequence data was available on GenBank for the plant species of interest or for closely related species. Primers were designed using Primer3Plus [21] to amplify regions of the two plastid genes (*matK*, *rbcL*) from which regions have been selected as barcode markers for land plants [22], two mitochondrial genes (*coxII*, *nad5*), and six low-copy nuclear genes (*ADH*-like, *EF1A*, *H3*, *hsp90*, *RD19*-like, *SKP1*), as well as the 18S rDNA gene (Table S2). These (high-quality and full-length) reference sequences were used for the subsequent design of nested primers for use in real-time PCR assays and 454-sequencing (see below).

PCR amplification was carried out in a 25- $\mu\text{l}$  reaction mixture that contained 10 to 50 ng of total DNA isolated from fresh plant material, 1 $\times$  DreamTaq<sup>TM</sup> buffer (Fermentas), 20  $\mu\text{g}$  BSA, 0.2 mM of each deoxynucleoside triphosphate (Promega), 10  $\mu\text{M}$  of each primer (Biolegio), and 1.0 U of DreamTaq<sup>TM</sup> DNA polymerase (Fermentas). The following thermocycling pattern was used to amplify each gene fragment: 94°C for 5 min (1 cycle); 94°C for 30 s, 55°C for 30 s and 72°C for 60 s (35 cycles); and then 72°C for 8 min (1 cycle). PCR products were purified using the Qiaquick PCR purification kit (Qiagen) following manufacturer's instructions. In order to verify sequence identity, purified PCR products were sequenced in both directions on an ABI9600 sequencer (Greenomics).

For *H3* and *SKP1* of *Laburnum anagyroides*, *ADH*-like of *Liriodendron tulipifera*, and *SKP1* of *Lonicera maackii* PCR products were cloned because PCR yields were insufficient for direct sequencing or because non-specific amplification products were detected. Clones were generated in *E. coli* using pGEM<sup>®</sup>-T Easy

Vectors (Promega). Sequences were submitted to EMBL under accession numbers FR869989–FR870021. Reference gene (sub-) families, cellular component ontology, and gene family size of target regions were identified using BLAST searches against the PLAZA 2.0 platform for plant comparative genomics database at <http://bioinformatics.psb.ugent.be/plaza/> [23].

### Real-time quantitative PCR (qPCR)

DNA from fresh and herbarium samples was subjected to qPCR assays targeting two plastid genes (*matK*, *rbcL*), two mitochondrial genes (*coxII*, *nad5*), and three of four low-copy nuclear genes (*ADH*-like, *EF1A*, *hsp90*, *SKP1*). PCR primers for real-time PCR assays, designed to amplify fragments between 88 to 140 bp, are reported in Table S3. Real-time detection of PCR products was conducted with SYBR Green I with the MyiQ detection system (Bio-Rad) and was conducted in a total volume of 20  $\mu\text{l}$ , containing 1 $\times$  iQ SYBR Green Supermix (Bio-Rad), 5  $\mu\text{M}$  of both primers, 20  $\mu\text{g}$  BSA, and 5 to 10 ng of total DNA. The temperature profile was: 95°C for 5 min (1 cycle); 95°C for 15 s, 60°C for 20 s, and 72°C for 20 s (40 cycles), with a final extension step at 72°C for 5 min. After amplification, melting curve analysis was performed from 60°C to 95°C, with increments of 0.5°C per 10 seconds. Real-time PCR standards were prepared by cloning target amplicons and subsequent amplification with M13 forward and reverse primers. The cloned PCR fragments were then purified and the yield was quantified using a NanoDrop 1000 (Thermo Scientific). Samples were analysed in triplicate (replicates of same DNA), and serial-diluted samples and no-template controls were included. Gene copy numbers were quantified absolutely by quantifying them per nanogram of total DNA. Amplification efficiencies for optimized qPCR assays calculated from slopes of the standard curves are shown in Table S3: efficiencies were between 1.90 and 2.09, with  $R^2$ -values, a measure for the quality of the regression fit, between 0.92 and 0.99. An exogenous control sequence from the wingless gene (*Wg*) of *Cymothoe Hüber* 1819 sp. (Lepidoptera, Insecta) was spiked to test for inhibition of qPCR assays by substances in the plant DNA extracts, but no inhibition was observed (data not shown). Gene copy numbers were considered to be duplicate (e.g., *rbcL* and *matK*, or *nad5* and *cox2*) or triplicate measurements (*ADH*-like, *EF1A*, and *hsp90* in *Liriodendron tulipifera*) in one-way ANOVA analyses, performed in PASW Statistics version 18.0.2 (SPSS Inc.). Copy number fold-reductions of plastid, mitochondrial, and nuclear genes directly after specimen preparation were expressed as 'fresh tissue/young herbarium gene copy number ratio', whereas those after long-term storage were expressed as 'young herbarium/old herbarium gene copy number ratio'.

### 454-sequencing of amplicons

Fusion primers for unidirectional 454 sequencing of amplicons of two plastid genes (*matK*, *rbcL*), two mitochondrial genes (*coxII*, *nad5*), six low-copy nuclear genes (*ADH*-like, *EF1A*, *H3*, *hsp90*, *RD19*-like, *SKP1*) and the 18S rDNA gene were constructed that incorporated the GS FLX Titanium forward primer A and a 10-bp multiplex identifier (MID) or reverse primer B (Table S4). The forward primer (Primer A-key) was: 5'-cctatcctatcctcggctgtctccgactcag-MID-template-specific-sequence-3'; and the reverse primer (Primer B-key) was 5'-cctatcccctgtgtcctggcagctctcag-template-specific-sequence-3'. MIDs 1 to 4 from the standard 454 set (Roche Technical Bulletin No. 013-2009) were selected, and each of the four species used was marked using a specific MID.

PCR amplification was conducted in a total volume of 50  $\mu\text{l}$ , containing 1 $\times$  High Fidelity PCR Buffer (Invitrogen), 5 mM of each deoxynucleoside triphosphate (Promega), 10  $\mu\text{M}$  of each primer (Biolegio), 50 mM  $\text{MgSO}_4$ , 1.0 U of Platinum<sup>®</sup>Taq High

Fidelity, 20 µg BSA and 5 to 10 ng of total DNA using the following thermo profile: 94°C for 1 min (1 cycle); 94°C for 30 s, 55°C for 30 s and 68°C for 60 s (35 cycles); and then 68°C for 8 min (1 cycle).

Amplicons were purified using the Agencourt AMPure XP kit (Beckman Coulter) and quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). Products from the same treatment (i.e. 'fresh tissue', 'young herbarium', or 'old herbarium') were mixed at equimolar concentrations to ca.  $1 \times 10^7$  molecules/µl. For each amplicon pool emPCR was performed following the emPCR Method Manual Lib-L for medium volume (Roche, 2009) and sequenced in a quarter of a PicoTiterPlate™ using a GS FLX Titanium Series. Pyrosequencing was carried out with primer A using the Genome Sequencer FLX (Roche Life Sciences Technology) at the Natural History Museum, University of Copenhagen, Denmark.

Amplicon sequence analysis was performed using the Galaxy platform [24,25] and included the following: first, a plot reflecting the base quality distribution was constructed. For each treatment, raw 454 sequence data and quality scores were converted to FASTQ format [26]. Low quality 3-prime ends were trimmed using a fixed value of 5% of the sequence length. Clean reads were reconverted to FASTA format and filtered by MID identifiers, constructing one file for each species-treatment.

Subsequently, reads were mapped against a reference sequence under the Roche-454 98% identity criterion and only reporting matches above 90% of identity and covering at least 50% of the reference sequence. A BAM file was generated from the SAM output, to only report bases with at least 100× coverage. The filtered pile up was examined for regions with a drastic drop in coverage, such as at the end of a homopolymer or at the end of the sequence, and these regions were manually removed from the output.

In sequences of *EF1A*, *hsp90*, *H3*, and *SKP1* a minority of nucleotide positions were found to be polymorphic in fresh controls and were omitted from further analysis. We assumed these to result from either allelic variation or to be paralogous sequence variants (not shown).

Nucleotide substitutions were counted across reads and grouped into six substitution types that are effectively indistinguishable in PCR reactions [27]; (A→C/T→G), (A→G/T→C), (A→T/T→A), (C→A/G→T), (C→G/G→C) and (C→T/G→A). The numbers of observed substitutions among the herbarium data was scaled to match the total amount of nucleotides in data from fresh tissue. For example, the corrected young herbarium count of plastid data for substitution pair (A→G/T→C) was calculated as: (Observed young herbarium plastid A→G/T→C)/(Total nucleotides in data fresh tissue plastid DNA)/(Total nucleotides in data young herbarium plastid DNA) =  $(11634 \times 10746437) / 14193327 = 8808.65$ . Chi-square ( $\chi^2$ ) tests of independence (at  $p \geq 0.05$  level) were used to investigate whether the distributions of

substitutions over the six substitution types were the same in sequence data obtained from fresh and herbarium DNA. The test was made two-sided because *a priori* the substitution direction from fresh DNA is not known. One-way analyses of variance (ANOVA) were performed to identify which of the six nucleotide substitution types occur at highest rates in each genomic compartment in herbarium DNA. For ANOVA, substitutions counts were expressed as numbers of substitutions per million total nucleotides observed (across reads).

## Supporting Information

**Figure S1** DNA extraction gel images. (DOCX)

**Table S1** Plant specimens, specimen information and mean DNA yield. (DOCX)

**Table S2** PCR primers and corresponding amplification targets. (DOCX)

**Table S3** Real-time PCR primers and performance of standard curves. (DOCX)

**Table S4** Titanium fusion primers used for 454-sequencing. (DOCX)

**Table S5** Copy numbers of chloroplast, mitochondrial and nuclear gene amplicons in fresh and herbarium tissues. (DOCX)

**Table S6** Mean DNA yield and mean gene copy numbers for plastid, mitochondrial and nuclear DNA regions. (DOCX)

**Table S7** ANOVA on average number of miscoding lesions (per  $10^6$  total nucleotides) for each substitution type across fresh and herbarium DNA. Separate analyses were performed for plastid, mitochondrial and nuclear DNA. (DOCX)

## Acknowledgments

The authors would like to thank Jan de Koning for his help in finding suitable herbarium specimens. Gerben Bijl is acknowledged for technical assistance.

## Author Contributions

Conceived and designed the experiments: MS FTB. Performed the experiments: MS AC RVG. Analyzed the data: MS AC. Contributed reagents/materials/analysis tools: MS AC GP OS. Wrote the paper: MS AC JER FTB.

## References

- Bebber DP, Carine MA, Wood JRI, Wortley AH, Harris DJ, et al. (2010) Herbaria are a major frontier for species discovery. *Proc Natl Acad Sci USA* 107: 22169–22171.
- Sebastian P, Schaefer H, Telford IRH, Renner SS (2010) Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. *Proc Natl Acad Sci USA* 107: 14269–14273.
- Metzker ML (2010) Sequencing technologies – the next generation. *Nat Rev Genet* 11: 31–46.
- Cubero OF, Crespo A, Fatchi J, Bridge PD (1998) DNA extraction and PCR amplification method suitable for fresh, herbarium-stored, lichenized, and other fungi. *Pl Syst Evol* 216: 243–249.
- Drábková L, Kirschner J, Vlcek C (2002) Comparison of seven DNA extraction and amplification protocols in historic herbarium specimens of Juncaceae. *Pl Mol Biol Rep* 20: 161–175.
- Telle S, Thines M (2008) Amplification of *cox2* (~620 bp) from 2 mg of up to 129 years old herbarium specimens, comparing 19 extraction methods and 15 polymerases. *PLoS One* 3: e3584.
- Doyle JJ, Dickson EE (1987) Preservation of plant species for DNA restriction endonuclease analysis. *Taxon* 36: 715–722.
- Pyle MM, Adams RP (1989) In situ preservation of DNA in plant specimens. *Taxon* 38: 576–581.
- Harris SA (1993) DNA analysis of tropical plant species: an assessment of different drying methods. *Pl Syst Evol* 188: 57–64.
- Chase MW, Hills HH (1991) Silica gel: an ideal material for field preservation of leaf samples for DNA studies. *Taxon* 40: 215–220.
- Srinivansan M, Sedmak D, Jewell S (2002) Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol* 161: 1961–1971.



12. Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* 362: 709–715.
13. Pääbo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, et al. (2004) Genetic analyses from ancient DNA. *Annu Rev Genet* 38: 645–679.
14. Lindahl T, Wood RD (1999) Quality control by DNA repair. *Science* 286: 1897–1905.
15. Savolainen V, Cuénoud P, Spichiger R, Martínez MDP, Crèvecoeur M, et al. (1995) The use of herbarium specimens in DNA phylogenetics: evaluation and improvement. *Pl Syst Evol* 197: 87–98.
16. Roldán-Arjona T, Ariza RR (2009) Repair and tolerance of oxidative DNA damage in plants. *Mut Res* 681: 169–179.
17. Boesch P, Weber-Lotfi F, Ibrahim N, Tarasenko V, Cosset A, et al. (2011) DNA repair in organelles: pathways, organization, regulation, relevance in disease and aging. *Biochim Biophys Acta* 1813: 186–200.
18. Gilbert MTP, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, et al. (2003) Distribution patterns of postmortem damage in human mitochondrial DNA. *Am J Hum Genet* 72: 48–61.
19. Stiller M, Green RE, Ronan M, Simons JF, Du L, et al. (2006) Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proc Natl Acad Sci USA* 103: 13578–13584.
20. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phyt Bull* 19: 11–15.
21. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, et al. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35: W71–W74.
22. Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratmasingham S, et al. (2009) A DNA barcode for land plants. *Proc Natl Acad Sci USA* 106: 12794–12797.
23. Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, et al. (2009) PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Pl Cell* 21: 3718–3731.
24. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, et al. (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* 19: 10.1–10.21.
25. Goecks J, Nekrutenko A, Taylor J, The Galaxy Team (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 25: R86.
26. Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, et al. (2010) Manipulation of FASTQ data with Galaxy. *Bioinform* 26: 1783–1785.
27. Hansen AJ, Willerslev E, Wiuf C, Mourier T, Arctander P (2001) Statistical evidence for miscoding lesions in ancient DNA templates. *Mol Biol Evol* 18: 262–265.
28. Pääbo S, Irwin DM, Wilson AC (1990) DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem* 265: 4718–4721.
29. Jobba LN, Roberts DL, Pimm SL (2011) How many species of flowering plants are there? *Proc Biol Sci* 278: 554–559.
30. Reape TJ, Molony EM, McCabe PF (2008) Programmed cell death in plants: distinguishing between different modes. *J Exp Bot* 59: 435–444.
31. McCabe PF, Levine A, Meijer PJ, Tapon NA, Pennell RI (1997) A programmed cell death pathway activated in carrot cells cultured at low cell density. *Plant J* 12: 267–280.
32. Gill SS, Tuteja N (2010) Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol Biochem* 48: 909–930.
33. Lindahl T, Nyberg B (1972) Rate of depurination of native deoxyribonucleic acid. *Biochem J* 11: 3610–3618.
34. Lindahl T (1990) Repair of intrinsic DNA lesions. *Mutation Res* 238: 305–311.
35. Hansen AJ, Mitchell DL, Wiuf C, Paniker L, Brand TB, et al. (2006) Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* 173: 1175–1179.
36. Heyn P, Stenzel U, Briggs AW, Kircher M, Hofreiter M, et al. (2010) Road blocks on paleogenomes – polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. *Nucleic Acids Res* 38: e161.
37. Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, et al. (2010) Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res* 38: e87.
38. Shapiro B (2008) Engineered polymerases amplify the potential of ancient DNA. *Trends Biotechnol* 26: 285–287.
39. Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, et al. (2007) Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of *post mortem* miscoding lesions. *Nucleic Acids Res* 35: 5717–5728.
40. Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, et al. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci USA* 104: 14616–14621.
41. Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res* 29: 4793–4799.
42. Smith DR (2009) Unparalleled GC content in the plastid DNA of *Selaginella*. *Plant Mol Biol* 71: 627–639.
43. Rowan BA, Oldenburg DJ, Bendich AJ (2009) A multiple-method approach reveals a declining amount of chloroplast DNA during development in *Arabidopsis*. *BCM Plant Biol* 9: 3.
44. Millar CD, Huynen L, Subramanian S, Mohandesan E, Lambert DM (2008) New developments in ancient genomics. *Trends Ecol Evol* 7: 386–393.